

La secuenciación de ADN mediante nanoporos y el ensamblaje *de novo*

Autor: Rinardo Belmar, Víctor Alfonso (Graduado en Biología, Máster en Microbiología, Alumno Predoctoral).

Público: Microbiólogos. **Materia:** Microbiología y genética. **Idioma:** Español.

Título: La secuenciación de ADN mediante nanoporos y el ensamblaje *de novo*.

Resumen

La Genómica es una disciplina de gran importancia actualmente. Su aplicación en campos de estudio dispares ha favorecido el desarrollo constante de nuevas tecnologías de secuenciación. Estas constituyen herramientas fundamentales para el estudio de los genomas. La secuenciación de ADN mediante nanoporos es una de ellas. La portabilidad y bajo coste de los secuenciadores basados en esta tecnología, así como su capacidad para determinar la secuencia de fragmentos de ADN de longitud mayor, la convierten en una de las técnicas de secuenciación más prometedoras. El objetivo de este trabajo es explicar esta tecnología y presentar sus principales ventajas e inconvenientes.

Palabras clave: ensamblaje *de novo*, genómica, nanoporos, secuenciación.

Title: DNA sequencing using nanopores and *de novo* assembly.

Abstract

Genomics is a discipline of great importance nowadays. Its application in disparate fields of study has favored the constant development of new sequencing technologies. These are fundamental tools for the study of genomes. The sequencing of DNA by nanopores is one of them. The portability and low cost of the sequencers based on this technology, as well as its capacity to determine the sequence of DNA fragments of greater length, make it one of the most promising sequencing techniques. The objective of this work is to explain this technology and present its main advantages and disadvantages.

Keywords: *de novo* assembly, genomics, nanopores, sequencing.

Recibido 2018-05-31; Aceptado 2018-06-05; Publicado 2018-06-25; Código PD: 096163

1. INTRODUCCIÓN

La Biología, como otras Ciencias, ha evolucionado con el tiempo. Uno de los avances más destacados del pasado siglo fue el comienzo de la Genómica. En la década de 1990, se desarrolló la tecnología necesaria para determinar el código genético completo de un organismo. Este hecho ha favorecido un cambio gradual en los métodos de estudio de las ciencias biológicas, desde una visión reduccionista hasta una perspectiva holística e integradora (Ellegren, 2014).

“La Genómica se ocupa de la caracterización molecular de genomas completos” (Griffiths, 2008). El análisis genómico hace uso de varias técnicas distintas, cuyo fin radica en el mejor conocimiento del origen, contenido, funcionamiento y evolución de los genomas. Una de estas técnicas es la secuenciación de genomas (Pierce, 2008).

Desde que Watson y Crick (1953) resolvieron la estructura tridimensional del ADN, conocer el orden de los nucleótidos de las cadenas de ácidos nucleicos ha supuesto un reto para la comunidad científica (Heather & Chain, 2016). Fue en 1975, cuando Sanger y Coulson (1975) describió por primera vez un método de secuenciación, la secuenciación enzimática. Dos años después, Maxam y Gilbert desarrollaron una técnica de secuenciación basada en otro principio: la secuenciación química (Maxam & Gilbert, 1977).

Las mejoras progresivas de las técnicas de secuenciación han dado lugar a distintas generaciones de secuenciación. Estas tecnologías han evolucionado de forma paralela al desarrollo del software necesario para el manejo de los datos obtenidos por las mismas (Heather & Chain, 2016).

Las tecnologías de la primera y segunda generación hicieron posible completar ambiciosos programas de secuenciación como son el Proyecto Genoma Humano y el Proyecto 1000 Genomas. No obstante, estos métodos solo permiten determinar pequeñas secuencias de ADN, lo que supone un inconveniente a la hora de ensamblar genomas *de novo*. Por tanto, existe una demanda de nuevas tecnologías que operen a una mayor velocidad y produzcan lecturas de secuencia de mayor longitud. Surge así la Tercera Generación de la Secuenciación (Lu, Giordano & Ning, 2016).

Las tecnologías de Tercera Generación, a diferencia de las otras, secuencian directamente moléculas individuales de ADN a tiempo real (Lu, Giordano, & Ning, 2016; Heather & Chain, 2016). Respecto a las anteriores generaciones, destacan tres mejoras: el aumento de la longitud de cada lectura desde 10 pares de bases hasta 104-106 pares de bases por lectura (Feng, Zhang, Ying, Wang & Du, 2015); la reducción del tiempo de secuenciación desde días a minutos y la ausencia del paso de amplificación por PCR. A esta generación pertenecen, PacBio (Pacific Biosciences) y la secuenciación de ADN mediante nanoporos (Oxford Nanopore Technology), la tecnología de secuenciación en la que se centra el presente trabajo (Laver, Harrison, O'Neill, Moore, Farbos, Paszkiewicz & Studholme, 2015).

La secuenciación de nanoporos no constituye una tecnología de desarrollo reciente; no obstante, presenta un gran potencial y es especialmente útil para el ensamblaje de genomas de novo. El ensamblaje de novo ofrece la posibilidad de conocer el genoma de un organismo sin disponer de referencias previas, permitiendo simplificar el proceso. Esta información contribuye al mejor conocimiento de organismos involucrados en procesos propios de campos tan diversos como son la medicina, industria y la agricultura. Por este motivo, los objetivos de este trabajo son describir y discutir el fundamento y las características de la secuenciación de ADN mediante nanoporos y presentar un procedimiento de ensamblaje de un genoma de novo.

2. SECUENCIACIÓN DE ADN MEDIANTE NANOPOROS

2.1. Fundamento teórico y características

La tecnología de nanoporos, que incluye a la secuenciación de ADN mediante nanoporos, se basa en el contador Coulter y los canales iónicos. Originalmente, la tecnología consistía en una cámara Coulter en la que los microcanales eran canales iónicos que permanecían abiertos gracias a la aplicación de un voltaje externo. Debido a este, las partículas con un tamaño ligeramente inferior al del tamaño del poro son capaces de atravesarlo. Los poros tienen un diámetro interno de un nanómetro y están o bien embebidos en membranas biológicas o bien son realizados sobre una película sólida de un material inorgánico. Tanto la membrana como la película inerte actúan como separadores entre las dos cámaras que contienen un fluido conductor. Los compartimentos se denominan cis y trans y en cada uno de ellos se sumerge un electrodo (Feng, 2015).

Bajo un voltaje, la solución de electrolitos se mueve a través del poro electroforéticamente y genera una corriente iónica que se puede medir como señal. Cuando el poro es bloqueado por un analito, como por ejemplo, una molécula de ADN que se añade en la cámara cis y que trata de pasar a la cámara trans a través del nanoporo, se interrumpe el flujo eléctrico. Las moléculas que bloquean el poro generan una resistencia al paso de la corriente eléctrica. Las variaciones en dicha resistencia permiten identificar la molécula (Feng, 2015; Lu, 2016).

Los nanoporos consisten específicamente en pequeños agujeros del orden de 1 nm de diámetro interno. Estos pueden ser divididos en dos categorías: biológicos y de estado sólido. Los nanoporos biológicos son proteínas transmembrana que conforman canales y que suelen estar embebidas en bicapas lipídicas, liposomas u otras películas de polímeros. Por otra parte, los nanoporos de estado sólido consisten en agujeros formados en superficies sintéticas como puede ser el SiO₂. Un caso particular de nanoporo de estado sólido es el nanoporo híbrido, el cual integra una proteína de poro en una superficie sintética (Feng, 2015).

La tecnología de nanoporos como herramienta que percibe moléculas únicas puede tener grandes aplicaciones en muchas áreas como el análisis de iones, ADN, ARN, proteínas, fármacos, polímeros y macromoléculas. El potencial de la tecnología de nanoporos aplicada a la secuenciación, se estableció en 1995 cuando se demostró que una hebra de ARN o ADN de cadena sencilla podía atravesar una bicapa lipídica a través de canales iónicos de α -hemolisina por electroforesis. El ADN o ARN sería leído nucleótido a nucleótido (Feng, 2015; Oxford Nanopore Technology, 2016).

En la actualidad se ha reconocido la utilidad de algunos nanoporos biológicos en secuenciación. Se conocen varios tipos de estos, como el de α -hemolisina. Este fue el primero que se desarrolló y es el más utilizado. Se trata de una exotoxina secretada por *Staphylococcus aureus*. Presenta una estructura en forma de champiñón, constituido por un heptámero con un canal con estructura de barril β . Forma un poro con 1.4 nm de diámetro interno mínimo, esta medida es suficiente para que una hebra de ADN de cadena sencilla (ADNcs) o de ARN pueda atravesar el canal. Como mejora, estos pueden ser modificados con un adaptador de ciclodextrina que contribuiría a regular el paso de la molécula de ADN a través del mismo (Feng, 2015; Bayley, 2015).

2.2. Secuenciadores automáticos

Los secuenciadores que existen en el mercado son comercializados por “Oxford Nanopore Technologies” (ONT), fundada por H. Bayley y G. Sanghera. MinION, el último sistema a la venta. Este dispositivo, del tamaño de un lápiz de memoria es capaz de realizar lecturas a tiempo real de 5.4 kb hasta 10 kb, muy superiores a las rendidas por el resto de tecnologías de secuenciación. El equipo tiene 512 nanoporos de α -hemolisina embebidos en un polímero sólido, lo que significa que puede leer 512 moléculas de ADN simultáneamente, aunque algunos nanoporos son más activos que otros (Bayley, 2015; Lu, 2016).

Al secuenciador se añade una muestra de ADN preparada especialmente para la secuenciación. Existen kits de preparación de las muestras suministrados por la misma casa comercial. En general, el método de preparación de la librería es similar al empleado por otras técnicas de secuenciación y consta de unas etapas básicas. En primer lugar, es necesario extraer el ADN, después se secciona en fragmentos de secuencia de más de 30 kb mediante centrifugación y se trata con un kit de reparación de errores. A cada fragmento de ADN se liga dos adaptadores, uno en cada extremo. El de la hebra líder se denomina “adaptador Y”, mientras que el adaptador con estructura de horquilla se llama “adaptador HP”. El primero, se une a una proteína motora y a una molécula de unión al nanoporo, mientras que, el segundo es un oligonucleótido con estructura de horquilla que es unido a una segunda proteína motora, denominada la proteína motora HP. El papel de las proteínas motoras es ayudar a que la molécula de ADN atraviese el poro e influyen en la velocidad con la que lo hace. (Lu, 2016; Reuter, Spacek & Snyder, 2015).

Cuando un fragmento de ADN de doble cadena se desnaturaliza, se obtienen las dos hebras complementarias de forma aislada. El “adaptador HP” conecta a las dos hebras complementarias a modo de horquilla, de forma que cuando la doble cadena se desnaturaliza, genera una molécula lineal de ADN constituida por una de las hebras, seguida del adaptador y la otra hebra complementaria. El siguiente paso es seleccionar los fragmentos de ADN del tamaño deseado y que tengan asociados un adaptador de cada tipo. Esto es posible gracias a

los propios adaptadores y a una columna de purificación de Histidina. Finalmente, al ADN se le adicionan las proteínas motoras, las cuales se anclan al fragmento ADN en la disolución, y una vez localizadas en la superficie del poro, favorece la apertura de la doble cadena y el paso de la hebra a través del poro a una velocidad que permite su detección por el sensor (Laver, 2015; Lu, 2016; Reuver, 2015). En el artículo que se ha analizado, utilizan el dispositivo MinION y para preparar las muestras utilizaron el protocolo estándar descrito anteriormente (Loman, Quick & Simpson, 2015).

La secuenciación comienza por el extremo 5' de cadena sencilla del “adaptador Y”, seguido de la cadena molde, luego se secuencian el “adaptador HP” y finalmente la cadena complementaria. A la información de una cadena se llama de una dirección (1D), y a la de dos cadenas, de dos direcciones (2D), es decir, la secuencia de la hebra líder y la de su complementaria. Durante el experimento, cada nanoporo sobre el chip sensor del array analiza moléculas de la muestra de forma independiente al resto de nanoporos. El análisis de datos se produce a tiempo real (Oxford Nanopore Technology, 2016; Laver, 2015; Lu, 2016; Reuver, 2015).

2.3 Ensamblaje de genomas

Una vez que el secuenciador genera los datos de secuenciación, es decir, la lectura y descodificación de los nucleótidos de DNA que componen el genoma del organismo correspondiente, la información recopilada es enviada al ensamblador para que lleve a cabo el ensamblaje completo (Miller, Koren & Sutton, 2010).

El proceso de ensamblaje consta de tres elementos jerárquicamente organizados. La lectura proporcionada por el secuenciador constituye el nivel jerárquico más bajo. Los contigs corresponden al siguiente nivel de jerarquía. Se puede definir como el alineamiento de múltiples lecturas que carecen de espacios y de orden definido. Finalmente, el scaffold ocupa el nivel más alto de la jerarquía del proceso. Se trata de la suma de dos o más contigs de forma ordenada y con la misma orientación que el genoma original. Definidos estos elementos, el proceso de ensamblaje parte de un conjunto de fragmentos de secuencia y trata de encontrar la “supersecuencia” común o secuencia origen de los fragmentos más corta, calcular alineamientos por pares de los fragmentos, seleccionar los dos fragmentos con el solapamiento mayor, mezclarlos y repetir este proceso hasta que se obtenga un solo fragmento de secuencia correspondiente al genoma completo alineado correctamente. Es un proceso que va “ensamblando” los elementos menores de la jerarquía, las lecturas, en contigs, y posteriormente estos contigs en scaffolds. El resultado de los ensambladores se puede medir por el tamaño y la precisión de sus contigs y scaffolds (Miller, Koren & Sutton, 2010).

Los tipos de ensamblaje de genomas pueden ser clasificados en dos grandes clases: Ensamblajes de novo y ensamblajes comparativos o Mapping. El ensamblaje comparativo o Mapping, consiste en el ensamblaje de un genoma a partir de una secuencia preexistente que es similar a la que se quiere ensamblar. El procedimiento básico coloca cada una de las lecturas en la posición adecuada utilizando el genoma de referencia como guía. Por el contrario, el ensamblaje de novo permite ensamblar largas secuencias sin disponer de ningún tipo de conocimiento previo acerca del genoma a ensamblar. Este tipo de ensamblajes son más lentos y presentan un mayor gasto computacional. Los ensambladores de novo utilizan algoritmos matemáticos de gran complejidad para generar los contigs necesarios para ensamblar la secuencia completa de un genoma (Vincent, Derome, Boyle, Culley & Charette, 2016). Uno de ellos es el ensamblador Overlap-Layout-Consensus (OLC) o en cadena, utilizado en el estudio de Loman, Quick y Simpson (2015) para ensamblar de novo el genoma de *E. coli*.

Los algoritmos basados en OLC han sido ampliamente utilizados para ensamblar pequeños genomas, aunque algunas aplicaciones han sido optimizadas para ensamblar genomas más grandes como por ejemplo Celera Assembler, Arachne, CAP y PCAP.

La tecnología MinION permite obtener lecturas superiores a las 10 kb, gracias a la tecnología de secuenciación de nanoporos. Este hecho permite esclarecer elementos repetitivos y copias dentro del genoma, de forma que las lecturas son lo suficientemente largas como para ser usadas eficazmente en el ensamblaje de novo.

3. DISCUSIÓN

Los avances en Genómica han supuesto una mejora sustancial en campos de la ciencia tan diversos como son la medicina, la agricultura, la industria o la ecología entre otros. La constante evolución y mejora de los secuenciadores ha contribuido a que en la actualidad se puedan obtener lecturas de secuencia más largas, precisas, con menor trabajo y a menor coste. Actualmente coexisten diferentes tecnologías de secuenciación. La elección del método de secuenciación adecuado está condicionado por los objetivos de la investigación. Para el ensamblaje de genomas de novo, las tecnologías de secuenciación de Tercera Generación, entre las que se encuentra la secuenciación de ADN mediante nanoporos y Pacific Biosciences, pueden resultar útiles. No obstante, pese a que permiten obtener lecturas de secuencia más largas que las de Segunda Generación, aspecto que facilita el tratamiento de los datos, deben de ser optimizadas con el objetivo de rendir lecturas menos imprecisas y más fiables. Disponer de lecturas de mayor calidad y tamaño, favorece un ensamblaje del genoma de novo más preciso y correcto, al obtener un menor número de contigs fácilmente superponibles que evitarían ciertos problemas asociados a secuencias repetitivas. Tanto la secuenciación de ADN mediante nanoporos como PacBio, generan lecturas de secuencia con una tasa de error mayor que las tecnologías de Segunda Generación. La secuenciación de ADN mediante nanoporos biológicos de α -hemolisina, arroja tasas de error que oscilan entre el 30 %, frente al 13 % de PacBio (da Fonseca et al., 2016). Como se observa, ambos sistemas presentan una tasa de error significativa, siendo superior en los secuenciadores de nanoporos de α -hemolisina.

Atendiendo a la alta tasa de error de la secuenciación de nanoporos mediante α -hemolisina, se podría inferir erróneamente que esta tecnología de secuenciación atisba un futuro poco prometedor, pero la realidad es diferente si se contemplan las diversas ventajas que ofrece respecto a otras tecnologías de secuenciación: genera lecturas de secuencia de mayor longitud que los secuenciados por PacBio, no utiliza reactivos fluorescentes para identificar las bases y no requiere la amplificación de ADN previa por PCR. Al prescindir de la amplificación, se evitan introducir errores y sesgos en las secuencias asociados a esta etapa, con lo que se reduce la carga de trabajo y se obtienen resultados con más rapidez. Además, secuenciadores como MinION y otros futuros que se están desarrollando, de pequeño tamaño, portátiles y disponibles a bajo coste, hacen que la secuenciación de nanoporos presente un gran atractivo para su uso en genómica (Lu, Giordano, & Ning, 2016; Feng, 2015; Bayley, 2015). Se espera que la secuenciación de nanoporos sea más efectiva en un futuro y se alcancen tasas de error en las lecturas inferiores a las ofrecidas por PacBio, que actualmente es el sistema de secuenciación de Tercera Generación más utilizado. Los problemas asociados de forma específica a la secuenciación de nanoporos son principalmente dos. Por una parte, los nanoporos de α -hemolisina están muy definidos y son altamente estables pero el canal es demasiado largo como para distinguir nucleótidos individuales directamente. La tasa de error observada en las lecturas de los secuenciadores de nanoporos de α -hemolisina puede responder, entre otros motivos, a la dificultad de interpretación de las señales eléctricas cuando se produce la entrada simultánea de dos nucleótidos en el canal del nanoporo. Este hecho no ha pasado desapercibido para los desarrolladores y actualmente se ensayan nuevos tipos de nanoporos, como los MspA, los nanoporos en estado sólido de grafeno y otros materiales, así como los nanoporos híbridos. El objetivo de estos avances es mejorar la calidad de la secuenciación y aumentar la sensibilidad de detección de cada nucleótido de la secuencia mediante la optimización del tamaño del canal del poro (Gupta, 2016;

Derrington et al. 2010). Por otra parte, el segundo problema está relacionado con la velocidad de paso de la molécula de ADN a través del poro, si esta fuera constante, cada intervalo en la señal podría ser asignado a cada nucleótido de la secuencia de forma exacta, de forma que las regiones de homopolímeros (AAAA ó TTTT) quedarían totalmente resueltas, sin embargo, esto no ocurre así; no se pueden asignar patrones de señal de cada nucleótido y como consecuencia no se pueden contabilizar de forma exacta cuantos nucleótidos del mismo tipo hay en la cadena. Este aspecto puede verse resuelto mediante la búsqueda de nuevas proteínas motoras (Loman, Quick & Simpson, 2015).

El proceso ensamblaje de genomas de forma general presenta una serie de inconvenientes que pueden conllevar a resultados erróneos. Los problemas vienen definidos por el número de secuencias repetidas que puede contener un genoma en su interior, que aproximadamente constituyen un 30 % (Li et al., 2012). Estas secuencias hacen más difícil el proceso de ensamblaje porque solapan en muchas regiones diferentes y es difícil determinar su emplazamiento inequívoco, debido a ello aumenta también la probabilidad de aparición “indels”. La posibilidad de obtener lecturas largas reduce el número de errores asociados a estos problemas.

Dado el alto porcentaje de error de la secuenciación mediante nanoporos, como estrategia adicional, se pueden realizar varias rondas de secuenciación, esto favorece posicionar de manera correcta cada secuencia al partir de un mayor número de estas. De hecho, dados los resultados que obtuvieron, los investigadores abogan por el empleo de datos del sistema MinION para ensamblar genomas de novo de forma precisa y con la orientación adecuada sin necesidad de utilizar datos de otros secuenciadores. No obstante, reconocen las deficiencias de esta tecnología que esperan que también puedan ser parcialmente corregidas con modelos probabilísticos futuros que incorporen información de la duración de translocación del fragmento que reviertan en una mayor precisión.

El sistema MinION puede tener múltiples aplicaciones científicas. De todas ellas, su uso para el ensamblaje de genomas de novo, es la aplicación con mayor interés. Otra aplicación es en el análisis de las muestras ambientales por metagenómica. De hecho, se ha utilizado para detectar nuevas especies de ranas en los bosques de Tanzania. Aunque el sistema debe ser mejorado, constituye una herramienta muy eficaz y con gran potencial para el análisis de la biodiversidad en investigaciones futuras. (Dodsworth, 2015). En el campo del diagnóstico y la medicina, tiene numerosas aplicaciones. Por citar alguna, en el campo de las enfermedades infecciosas, permite facilitar la identificación del microorganismo implicado en la enfermedad y determinar cuál va a ser la evolución viral, investigar la transmisión de cadenas o validar ensayos diagnósticos. Es especialmente útil ante una situación en que la velocidad en la identificación de la bacteria causante de la infección es crítico para la supervivencia del paciente. Por ejemplo, gracias a la capacidad del secuenciador MinION para analizar muestras *in situ*, ofrece oportunidades fantásticas para analizar genomas en condiciones menos favorables.

4. CONCLUSIÓN

En palabras de F. Sanger “[...] knowledge of sequences could contribute much to our understanding of living matter”, el conocimiento de las secuencias contribuye a nuestra forma de entender de la vida y es que a partir de esta información obtenida, la cantidad de aplicaciones es infinita, desde identificar genes que produzcan proteínas de interés biofarmacéutico o industrial hasta describir nuevas especies e incluso caracterizar totalmente su genoma.

Sin embargo, es imprescindible conocer el fundamento y alcance de estos métodos para poder realizar una interpretación objetiva y precisa de los datos obtenidos, no solo para lograr un conocimiento científico de calidad, sino para ser conscientes de los pasos que hay que seguir para seguir mejorando la tecnología actual. Los métodos de secuenciación no son perfectos y tienen limitaciones que hay que contrarrestar mediante el empleo de otras técnicas. Por ejemplo, la secuenciación de ADN a través de nanoporos sería una de las tecnologías más prometedoras pero debe mejorar sustancialmente, pues los resultados obtenidos con ella, revelan, porcentajes de error que serían inadmisibles en el caso de que se usase exclusivamente esta técnica. Asimismo, el gigantesco volumen de datos que hay que procesar requiere del desarrollo paralelo de potentes herramientas bioinformáticas intuitivas y que requieran un menor gasto computacional.

Para finalizar este trabajo, hay que señalar que la secuenciación de ADN mediante nanoporos, es útil, rápida, accesible y versátil. Sin embargo, aunque facilita la labor investigativa actual, no se deberían abandonar el resto de ensayos genéticos, pues son otra herramienta con la que mejorar los resultados obtenidos por esta técnica a la que aún le queda un amplio programa de desarrollo, tanto a nivel del fundamento mecánico, como bioinformático.

Bibliografía

- Bayley, H. (2015). Nanopore sequencing: from imagination to reality. *Clinical Chemistry*, 61(1), 25–31. DOI:10.1373/clinchem.2014.223016
- da Fonseca R.R., Albrechtsen A., Themudo G.E., Ramos-Madrigal J., Sibbesen J.A., Maretty L., Zepeda-Mendoza M.L., Campos P.F., Heller R. & Pereira R.J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, pii: S1874-7787(16), 30036-30038. DOI: 10.1016/j.margen.2016.04.012. [Epub ahead of print]
- Derrington, I.M., Butler, T. Z., Collins, M. D., Manrao E., Pavlenok, M., Niederweis M. & Gundlach J.H. (2010) Nanopore DNA sequencing with MspA. *Proceedings of the National Academy of Sciences of the United States of America*, 107(37), 16060-16065.
- Dodsworth S. (2015). Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science*, 20(9), 525-527. DOI: 10.1016/j.tplants.2015.06.012
- Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, 29(1), 51- 63. Doi: 10.1016/j.tree.2013.09.008
- Feng, Y., Zhang, Y., Ying, C., Wang, D. & Du, C. (2015). Nanopore-based Fourth-generation DNA Sequencing Technology. *Genomics, Proteomics & Bioinformatics*, 13(1) 4-16. DOI: 10.1016/j.gpb.2015.01.009
- Griffiths, A., Wessler, S., Lewontin, R. & Carroll, S. (2008). *Genética*. Madrid: McGrawHill
- Gupta P.D. (2016). Nanopore Technology: A Simple, Inexpensive, Futuristic Technology for DNA Sequencing. *Indian Journal of Clinical Biochemistry*, 31(4), 359-360. DOI: 10.1007/s12291-016-0593-6
- Heather, J.M. & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107, 1-8. Doi: 10.1016/j.ygeno.2015.11.003
- Laver, T., Harrison J., O'Neill P.A., Moore, K., Farbos, A., Paszkiewicz K. & Studholme, D.J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification*, 3, 1-8. DOI: 10.1016/j.bdq.2015.02.001
- Li Z., Chen, Y., Mu D., Yuan, J., Shi Y., Zhang, H, Gan, J., Li, N., Hu, X., Liu, B., Yang B. & Fan, W. (2012). Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Briefings in Functional Genomics*, 11(1), 25-37. DOI: 0.1093/bfpg/elr035
- Loman, N. J., Quick, J. & Simpson, J. T. (2015) A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods*, 12(8), 733-735. DOI: 10.1038/nmeth.3444
- Lu, H., Giordano, F. & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*. Doi: 10.1016/j.gpb.2016.05.004
- Maxam, A.M. & Gilbert W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), 560-564.
- Miller, J.R., Koren, S. & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, 95, 315-327. DOI:10.1016/j.ygeno.2010.03.001
- Pierce, B. A. (2008). *Genomics and Proteomics. Genetics: A conceptual Approach (3ra.ed)*, 548-552. W. H. Freeman & Company
- Reuter, J.A., Spacek, D. V. & Snyder, M.P. (2015) High-Throughput Sequencing Technologies. *Molecular Cell*, 58, 586-597. DOI: 10.1016/j.molcel.2015.05.004
- Sanger, F. & Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), 441-448. DOI: 10.1016/0022-2836(75)90213-2
- Vincent, A.T., Derome, N., Boyle, B., Culley, A.I. & Charette, S. J. (2016). Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *Journal of Microbiological Methods*. DOI: 10.1016/j.mimet.2016.02.01
- Watson, J.D. & Crick, F.H.C. (1953). Molecular Structure of Deoxyribose Nucleic Acids. *Nature*, 171, 737-738